

# Application of Information Theory, Lecture 1

## Basic Definitions and Facts

### Handout Mode

Iftach Haitner

Tel Aviv University.

October 28, 2014

# The entropy function

$X$  — Discrete random variable (finite number of values) over  $\mathcal{X}$  with probability mass  $p = p_X$ . The **entropy** of  $X$  is defined by:

$$H(X) := - \sum_{x \in \mathcal{X}} \Pr[X = x] \cdot \log_2 \Pr[X = x]$$

taking  $0 \log 0 = 0$ .

- ▶  $H(X) = - \sum_x p(x) \log p(x) = E_X \log \frac{1}{p(X)} = E_{Y=p(X)} \log \frac{1}{Y}$
- ▶  $H(X)$  was introduced by Shannon as a measure for the uncertainty in  $X$  — number of **bits** required to describe  $X$ , information we don't have about  $X$ .
- ▶ When using the natural logarithm, the quantity is called **nats** ("natural")
- ▶ Entropy is a function of  $p$  (sometimes refers to as  $H(p)$ ).

# Examples

1.  $X \sim (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ :

(i.e., for some  $x_1 \neq x_2 \neq x_3$ ,  $P_X(x_1) = \frac{1}{2}$ ,  $P_X(x_2) = \frac{1}{4}$ ,  $P_X(x_3) = \frac{1}{4}$ )

$$H(X) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{4} \log \frac{1}{4} = \frac{1}{2} + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 2 = 1\frac{1}{2}.$$

2.  $H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ .

3.  $X$  is uniformly distributed over  $\{0, 1\}^n$ :

$$H(X) = -\sum_{i=1}^{2^n} \frac{1}{2^n} \log \frac{1}{2^n} = -\log \frac{1}{2^n} = n.$$

▶  $n$  bits are needed to describe  $X$

▶  $n$  bits are needed to create  $X$

4.  $X = X_1, \dots, X_n$  where  $X_i$  are iid over  $\{0, 1\}^n$ , with  $P_X(1) := \Pr[X = 1] = \frac{1}{3}$ .  $H(X) = ?$

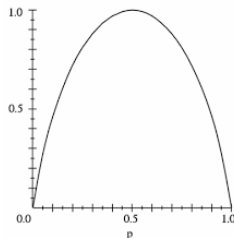
5.  $X \sim (p, q)$ ,  $p + q = 1$

▶  $H(X) = H(p, q) = -p \log p - q \log q$

▶  $H(1, 0) = (0, 1) = 0$

▶  $H(\frac{1}{2}, \frac{1}{2}) = 1$

▶  $h(p) := H(p, 1-p)$  is continuous



## Axiomatic derivation of the entropy function

Any other choices for defining entropy?

Shannon function is the **only** symmetric function (over probability distributions) satisfying the following three axioms:

**A1** Continuity:  $H(p, 1 - p)$  is continuous function of  $p$ .

**A2** Normalization:  $H(\frac{1}{2}, \frac{1}{2}) = 1$

**A3** Grouping axiom:

$$H(p_1, p_2, \dots, p_m) = H(p_1 + p_2, p_3, \dots, p_m) + (p_1 + p_2)H(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2})$$

Why **A3**?

Not hard to prove that Shannon's entropy function satisfies above axioms, proving this is the only such function is more challenging.

Let  $H$  be a function that satisfying the above axioms.

We prove (assuming additional axiom) that  $H$  is the Shannon function.

## Generalization of the grouping axiom

Fix  $p = (p_1, \dots, p_m)$  and let  $S_k = \sum_{i=1}^k p_i$ .

Grouping axiom:  $H(p_1, p_2, \dots, p_m) = H(S_2, p_3, \dots, p_m) + S_2 H(\frac{p_1}{S_2}, \frac{p_2}{S_2})$ .

### Claim 1 (Generalized grouping axiom)

$$H(p_1, p_2, \dots, p_m) = H(S_k, p_{k+1}, \dots, p_m) + S_k \cdot H(\frac{p_1}{S_k}, \dots, \frac{p_k}{S_k})$$

Proof: Let  $h(q) = H(q, 1 - q)$ .

$$\begin{aligned} H(p_1, p_2, \dots, p_m) &= H(S_2, p_2, \dots, p_m) + S_2 h(\frac{p_2}{S_2}) \\ &= H(S_3, p_3, \dots, p_m) + S_3 h(\frac{p_3}{S_3}) + S_2 h(\frac{p_2}{S_2}) \\ &\vdots \\ &= H(S_k, p_{k+1}, \dots, p_m) + \sum_{i=2}^k S_i h(\frac{p_i}{S_i}) \end{aligned} \tag{1}$$

Hence,

$$H(\frac{p_1}{S_k}, \dots, \frac{p_k}{S_k}) = H(\frac{S_{k-1}}{S_k}, \frac{p_k}{S_k}) + \sum_{i=1}^{k-1} \frac{S_i}{S_k} h(\frac{p_i/S_k}{S_i/S_k}) = \frac{1}{S_k} \sum_{i=1}^k S_i h(\frac{p_i}{S_i}) \tag{2}$$

Claim follows by combining the above equations.  $\square$

## Further generalization of the grouping axiom

Let  $1 = k_1 < k_2 < \dots < k_q < m$  and let  $C_t = \sum_{i=k_t}^{k_{t+1}-1} p_i$  (letting  $k_{q+1} = m + 1$ ).

### Claim 2 (Generalized<sup>++</sup> grouping axiom)

$$H(p_1, p_2, \dots, p_m) = \\ H(C_1, \dots, C_q) + C_1 \cdot H\left(\frac{p_1}{C_1}, \dots, \frac{p_{k_2-1}}{C_1}\right) + \dots + C_q \cdot H\left(\frac{p_{k_q+1}}{C_q}, \dots, \frac{p_m}{C_q}\right)$$

Proof: Follow by the extended group axiom and the symmetry of  $H$   $\square$

Implication: Let  $f(m) = H(\underbrace{\frac{1}{m}, \dots, \frac{1}{m}}_m)$

$$\triangleright f(3^2) = 2f(3) = 2H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

$$\implies f(3^n) = nf(3).$$

$$\triangleright f(mn) = f(m) + f(n)$$

$$\implies f(m^k) = kf(m)$$

$$f(m) = \log m$$

We give a proof under the additional axiom

$$\mathbf{A4} \quad f(m) < f(m+1)$$

(you can Google for a proof using only **A1–A3**)

- ▶ For  $n \in \mathbb{N}$  let  $k = \lfloor n \log 3 \rfloor$ .
- ▶ By **A4**,  $f(2^k) < f(3^n) < f(2^{k+1})$ .
- ▶ By grouping axiom,  $k < nf(3) < k+1$ .

$$\implies \frac{\lfloor n \log 3 \rfloor}{n} < f(3) < \frac{\lfloor n \log 3 \rfloor + 1}{n} \text{ for any } n \in \mathbb{N}$$

$$\implies f(3) = \log 3.$$

- ▶ Proof extends to any integer (not only 3)

$$H(p, q) = -p \log p - q \log q$$

- ▶ For **rational**  $p, q$ , let  $p = \frac{k}{m}$  and  $q = \frac{m-k}{m}$ , where  $m$  is the smallest common multiplier.
- ▶ By grouping axiom,  $f(m) = H(p, q) + p \cdot f(k) + q \cdot f(m - k)$ .
- ▶ Hence,

$$\begin{aligned} H(p, q) &= \log m - p \log k - q \log(m - k) \\ &= p(\log m - \log k) + q(\log m - \log(m - k)) \\ &= -p \log \frac{m}{k} - q \log \frac{m - k}{m} = -p \log p - q \log q \end{aligned}$$

- ▶ By continuity axiom, holds for **every**  $p, q$ .



$$H(p_1, p_2, \dots, p_m) = - \sum_i^m -p_i \log p_i$$

We prove for  $m = 3$ . Proof for arbitrary  $m$  follows the same lines.

- ▶ For rational  $p_1, p_2, p_3$ , let  $p_1 = \frac{k_1}{m}$ ,  $p_2 = \frac{k_2}{m}$  and  $p_3 = \frac{k_3}{m}$ , where  $m = k_1 + k_2 + k_3$  is the smallest common multiplier.
- ▶  $f(m) = H(p_1, p_2, p_3) + p_1 f(k_1) + p_2 f(k_2) + p_3 f(k_3)$
- ▶ Hence,

$$\begin{aligned} H(p_1, p_2, p_3) &= \log m - p_1 \log k_1 - p_2 \log k_2 - p_3 \log k_3 \\ &= -p_1 \log \frac{k_1}{m} - p_2 \log \frac{k_2}{m} - p_3 \log \frac{k_3}{m} \\ &= -p_1 \log p_1 - p_2 \log p_2 - p_3 \log p_3 \end{aligned}$$

- ▶ By continuity axiom, holds for every  $p_1, p_2, p_3$ .

$$0 \leq H(p_1, \dots, p_m) \leq \log m$$

► Tight bounds

- $H(p_1, \dots, p_m) = 0$  for  $(p_1, \dots, p_m) = (1, 0, \dots, 0)$ .
- $H(p_1, \dots, p_m) = \log m$  for  $(p_1, \dots, p_m) = (\frac{1}{m}, \dots, \frac{1}{m})$ .

► Non negativity is clear.

- A function  $f$  is **concave** if  $\forall t_1, t_2, \lambda \in [0, 1] \leq 1$   
 $\lambda f(t_1) + (1 - \lambda)f(t_2) \leq f(\lambda t_1 + (1 - \lambda)t_2)$

$$\Rightarrow \text{(by induction)} \quad \forall t_1, \dots, t_k, \lambda_1, \dots, \lambda_k \in [0, 1] \text{ with } \sum_i \lambda_i = 1$$

$$\sum_i \lambda_i f(\lambda_i t_i) \leq f(\sum_i \lambda_i t_i)$$

$$\Rightarrow \text{(Jensen inequality): } E f(X) \leq f(E X) \text{ for any random variable } X.$$

- $\log(x)$  is (strictly) concave for  $x > 0$ , since its second derivative  $(-\frac{1}{x^2})$  is always negative.
- Hence,  $H(p_1, \dots, p_m) = \sum_i p_i \log \frac{1}{p_i} \leq \log \sum_i p_i \frac{1}{p_i} = \log m$
- Alternatively, for  $X$  over  $\{1, \dots, m\}$ ,  
 $H(X) = E_X \log \frac{1}{P_X(X)} \leq \log E_X \frac{1}{P_X(X)} = \log m$

$$H(g(X)) \leq H(X)$$

Let  $X$  be a random variable, and let  $g$  be over  $\text{Supp}(X) := \{x: P_X(x) > 0\}$ .

►  $H(Y = g(X)) \leq H(X)$ .

Proof:

$$\begin{aligned} H(X) &= - \sum_x P_X(x) \log P_X(x) = - \sum_y \sum_{x: g(x)=y} P_X(x) \log P_X(x) \\ &\geq - \sum_y P_Y(y) \max_{x: g(x)=y} \log P_X(x) \\ &\geq - \sum_y P_Y(y) \log P_Y(y) = H(Y) \end{aligned}$$

► If  $g$  is injective, then  $H(Y) = H(X)$ .

Proof:  $p_X(X) = P_Y(Y)$ .

► If  $g$  is non-injective (over  $\text{Supp}(X)$ ), then  $H(Y) < H(X)$ .

Proof: ?

►  $H(X) = H(2^X)$ .

►  $H(X) < H(\cos(X))$ , if  $0, \pi \in \text{Supp}(X)$ .

# Notation

- ▶  $[n] = \{1, \dots, n\}$
- ▶  $P_X(x) = \Pr[X = x]$
- ▶  $\text{Supp}(X) := \{x : P_X(x) > 0\}$
- ▶ For random variable  $X$  over  $\mathcal{X}$ , let  $p(x)$  be its density function:  
 $p(x) = P_X(x)$ .  
In other words,  $X \sim p(x)$ .
- ▶ For random variable  $Y$  over  $\mathcal{Y}$ , let  $p(y)$  be its density function:  
 $p(y) = P_Y(y)$ ...